

<https://helda.helsinki.fi>

The Impact of Covid-19 on Smartphone Usage

Li, Tong

2021-12-01

Li , T , Zhang , M , Li , Y , Lagerspetz , E , Tarkoma , S & Hui , P 2021 , ' The Impact of Covid-19 on Smartphone Usage ' , IEEE internet of things journal , vol. 8 , no. 23 , pp. 16723-16733 . <https://doi.org/10.1109/JIOT.2021.3073864>

<http://hdl.handle.net/10138/336963>

<https://doi.org/10.1109/JIOT.2021.3073864>

unspecified

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

The Impact of Covid-19 on Smartphone Usage

Tong Li, *Student Member, IEEE*, Mingyang Zhang, Yong Li, *Senior Member, IEEE*,
Eemil Lagerspetz, Sasu Tarkoma, *Senior Member, IEEE*, and Pan Hui, *Fellow, IEEE*

Abstract—The outbreak of Covid-19 changed the world as well as human behavior. In this paper, we study the impact of Covid-19 on smartphone usage. We gather smartphone usage records from a global data collection platform called Carat, including the usage of mobile users in North America from November 2019 to April 2020. We then conduct the first study on the differences in smartphone usage across the outbreak of Covid-19. We discover that Covid-19 leads to a decrease in users’ smartphone engagement and network switches, but an increase in WiFi usage. Also, its outbreak causes new typical diurnal patterns of both memory usage and WiFi usage. Additionally, we investigate the correlations between smartphone usage and daily confirmed cases of Covid-19. The results reveal that memory usage, WiFi usage, and network switches of smartphones have significant correlations, whose absolute values of Pearson coefficients are greater than 0.8. Moreover, smartphone usage behavior has the strongest correlation with the Covid-19 cases occurring after it, which exhibits the potential of inferring outbreak status. By conducting extensive experiments, we demonstrate that for the inference of outbreak stages, both Macro-F1 and Micro-F1 can achieve over 0.8. Our findings explore the values of smartphone usage data for fighting against the epidemic.

Index Terms—Smartphone usage, Covid-19, correlations, outbreak stage inference.

I. INTRODUCTION

At the beginning of 2020, Covid-19 was identified and has spread globally [1]. The outbreak of Covid-19 has changed people’s lives significantly. Countless efforts have been made to study the world after Covid-19 from different perspectives, ranging from world economy [2], personal mental health [3], to human mobility [4]. Meanwhile, since the first iPhone was released in 2007, smartphones have become a necessity in daily lives [5]. The number of smartphone users worldwide today has surpassed three billion [6]. However, up to now, the understanding of the impact of Covid-19 on smartphone usage is still inadequate. Specifically, studying how Covid-19 affects users’ smartphone usage behavior can bring two-fold benefits. First, understanding smartphone usage differences across the Covid-19 outbreak is critical for the industry, e.g., smartphone manufacturers and network service providers, to dynamically adjust market strategies and enhance user experience. Second, smartphones are embedded with a set of sensors recording user

activities in both cyber and physical spaces [7]. By exploring the impact, we can use such rich behavioral data to infer different Covid-19 outbreak stages and further contribute to the fight against Covid-19.

Meanwhile, some previous studies have introduced mobile sensing data to the public health field. For example, Yarkoni [8] proposed the concept of Psychoinformatics, using tools and techniques from information sciences to improve psychological research. Insel [9] and Baumeister *et al.* [10] introduced digital phenotyping that leverages digital behavior data logged on smartphone sensors to detect psychological states. Further, Markowitz *et al.* [11] proposed to explore big data technologies and conduct digital phenotyping on a large-scale. The above studies showed the correlation between smartphone usage and the psychological states of users. The Covid-19 pandemic represents a global health crisis, which will severely change psychological burdens and physical activities of individuals [12]. Such changes may be conveyed to and reflected in smartphone usage [13]. In this way, we are motivated to investigate how the outbreak of Covid-19 affects smartphone usage behavior.

In this work, we make an effort towards understanding the impact of Covid-19 on smartphone usage and explore the potential of smartphone usage data to fight against Covid-19. More specifically, we study the following research problems.

- 1) Does the outbreak of Covid-19 affect users’ smartphone usage, and how?
- 2) Can we use smartphone usage data, e.g., CPU usage, memory usage, and network connections, to infer the outbreak stages of Covid-19?

To answer the above two questions, we reveal the correlations between smartphone usage and the outbreak of Covid-19 from both statistics and dynamic patterns. We first collect a large-scale smartphone usage dataset by leveraging a global crowdsourcing platform called Carat. The dataset covers users in North America and their smartphone usage records for six months from November 2019 to April 2020 (Section II). Next, we use the dataset to make a statistical analysis. The results demonstrate that the outbreak of Covid-19 has indeed impacted significantly on users’ smartphone usage behavior in terms of CPU usage, memory usage, WiFi usage, and network switches. In our case, CPU and memory usage describe how much of the processor’s and memory’s capacity is in use, respectively. WiFi usage indicates the percentage of records under WiFi connection. Network switch refers to the change of network connection from WiFi to cellular network and vice versa. Specifically, the CPU usage and memory usage reflect the intensity of smartphone engagement of users. The WiFi usage and network switches reveal users’ mobility intensity.

T. Li, M. Zhang and P. Hui are with the System and Media Laboratory (SyMLab), Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong. T. Li and P. Hui are also with the Department of Computer Science, University of Helsinki, Helsinki, Finland. (E-mail: t.li@connect.ust.hk, mzhangbj@ust.hk, panhui@cse.ust.hk)

Y. Li is with Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. (E-mail: liyong07@tsinghua.edu.cn)

E. Lagerspetz and S. Tarkoma are with the Department of Computer Science, University of Helsinki, Helsinki, Finland. (E-mail: eemil.lagerspetz@cs.helsinki.fi, sasu.tarkoma@helsinki.fi)

TABLE I
SAMPLES OF THE COLLECTED SMARTPHONE USAGE DATA, WHERE LTE AND UTMS ARE SPECIFIC MODES OF CELLULAR NETWORK. USER IDS HAVE BEEN ANONYMIZED.

User ID	Timestamp	CPU usage (%)	Active memory (KB)	Free memory (KB)	Network	Battery level (%)	Timezone	MCC
1	2019-11-05 05:48:48	0.6992	3528	1875820	WiFi	66	America/Denver	us
2	2019-11-13 07:51:11	51.5152	429632	1746900	LTE	79	America/New_York	us
3	2019-11-16 12:31:51	53.1807	1006248	2853892	UTMS	94	Europe/Helsinki	fi

Further, we extend our analysis to dynamic patterns, i.e., diurnal patterns of smartphone usage. The results unveil how the outbreak of Covid-19 affects usage behavior during the time of one day. We also examine the correlations between smartphone usage and daily confirmed cases (Section III). Moreover, we investigate smartphone usage data's inference ability for Covid-19 outbreak stages using both statistical and deep learning methods. By comparing the performance and conducting importance analysis, we select the most potent smartphone usage features for the outbreak stage inference (Section IV).

Among the many insightful results and observations, the following are the most prominent.

- The outbreak of Covid-19 causes a decrease in users' smartphone engagement in terms of both CPU usage and memory usage. However, it has different impacts on CPU and memory usage according to their diurnal patterns. Specifically, it leads to a new typical diurnal pattern of memory usage while it only changes the proportion of existing patterns of CPU usage.
- The outbreak of Covid-19 makes an increase in WiFi usage and a decrease in network switches, implying that users reduce their mobility intensity. Also, similar to memory usage, a new typical diurnal pattern of WiFi usage has emerged after the outbreak.
- Memory usage, WiFi usage, and network switches have significant correlations with the number of daily confirmed cases of Covid-19. Also, the correlation between smartphone usage behavior and Covid-19 daily cases has a time delay. Smartphone usage changes earlier than the number of cases. That is because the smartphone data can reflect the outbreak status in real-time. However, such reflection cannot be immediately expressed in daily cases due to the diagnosis delay.
- By using smartphone usage data to infer Covid-19 outbreak stages, we can achieve over 0.8 for both Macro-F1 and Micro-F1, which presents a promising application of smartphone usage data on fighting against Covid-19.

II. DATASET OVERVIEW

A. Data Collection

We leverage a crowdsourcing platform called Carat to collect smartphone usage data. Carat is a cross operating system mobile app, including both iOS¹ and Android², which can record users' smartphone usage traces automatically. Carat can monitor and record the working status of smartphones in

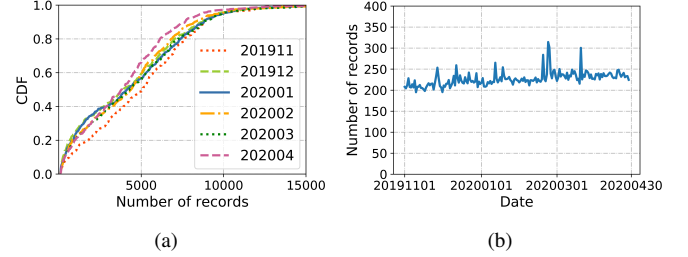


Fig. 1. This figure shows (a) The cumulative distribution function (CDF) of the number of records per month for each unique user. (b) The daily average number of collected records of users.

detail. In practice, Carat informs of all data collection items in the End-user License Agreement (EULA) when users install Carat to alleviate user privacy concerns. Also, Carat users are anonymized, and the app does not collect any personal information. It is worth noting that Carat is live. Up to now, Carat has been downloaded over 100 thousand times. The number of downloads and installations is increasing every day.

Specifically, Carat applies an event-triggered collection scheme, gathering a data sample every time the battery level changes by 1%. Each data sample contains a list of smartphone hardware status, including CPU, memory, battery, and network. Each sample also has several other features, including a user-specific identifier, timestamp, timezone, and mobile country code (MCC). The MCC is obtained from the cellular network and automatically converted to a two-character country code. Table I presents samples of collected smartphone usage data to show the data format.

B. Basic Analysis

Since we focus on studying the impact of Covid-19, we select the records from November 2019 to April 2020. Also, we principally consider samples collected from North America. In total, we have 452 users with over 7,517,494 records. Since users involved may uninstall and reinstall Carat during the data collection period, the number of active users changes over time, i.e., November 2019 (293 users), December 2019 (295 users), January 2020 (251 users), February 2020 (224 users), March 2020 (198 users), and April 2020 (158 users). In our case, we use both timezones and MCC to determine the users' country, which increases the reliability of detection. Table II summarizes the dataset.

Next, we depict basic statistics to illustrate the quality and representativeness of the collected smartphone usage dataset. Fig. 1(a) presents the cumulative distribution function (CDF) of the number of records per month for each unique user. We observe that the involved users kept a high activeness level

¹<https://apps.apple.com/us/app/carat/id504771500>

²<https://play.google.com/store/search?q=carat>

TABLE II
SUMMARY OF THE COLLECTED DATASET FROM THE USA.

# Users	# Records	Attributes	Date	Area
452	7,517,494	User ID, timestamp, CPU usage, Memory usage, network status, battery level, timezone, MCC	11/2019 - 04/2020	North America

during the data collection period. For each month, more than 20% of users have over 1,800 records. Moreover, we plot how the average number of users' records changes every day in Fig. 1(b). We can witness that there are around 220 records every day per user on average. Although there are some fluctuations, the curve is relatively stable. Such a high number of records per user demonstrates our dataset's effectiveness in capturing the smartphone usage behavior of users involved covering the entire six months, i.e., from November 2019 to April 2020. Also, the continuity of the data collection guarantees the representativeness of our study.

C. Ethical Considerations

We are very aware of the privacy issues when using the collected data for research. We have taken adequate actions to safeguard the privacy of the involved mobile users. **First**, we do not collect any personal information from users. A user-specific identifier is randomly generated when the user first installs Carat. We only have users' country information rather than sensitive location information, like GPS data. Also, the data-gathering part of Carat is open-source³. Users can examine it easily. The users involved are informed of the data collection and management procedures in the End-user License Agreement (EULA) and grant their consent from their devices. In the EULA, we also point out that the data we collect may be used to improve products or for research purposes. **Second**, the dataset is stored in a secure local server protected by strict authentication mechanisms and firewalls. All researchers are regulated by a strict non-disclosure agreement to access the data. **Finally**, this work has received approval from all the authors' local institutions.

III. DIFFERENCES IN SMARTPHONE USAGE

In this section, we aim to solve the first research problem, i.e., whether and how the outbreak of Covid-19 affects users' smartphone usage behavior. Specifically, we explore the impact on CPU usage, memory usage, and network status from statistical and dynamic pattern analysis. The data processing and analysis was conducted in Helsinki.

A. Differences in Number and Distributions

To determine whether the outbreak of Covid-19 changes users' mobile engagement, first of all, we need to determine the outbreak date in North America. Fig. 2 shows the cumulative number of confirmed cases in North America from February 2020 to April 2020 and the governmental policies on the same timescale. The dashed curve is in the linear scale, while the solid curve depicts the cumulative number

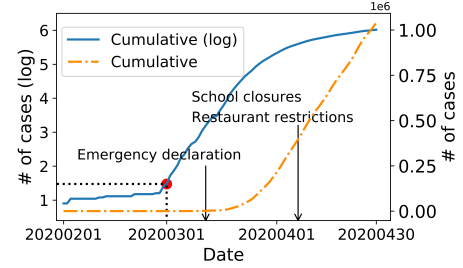


Fig. 2. The cumulative number of confirmed cases changes over time. The federal government issued an emergency declaration on March 13, 2020. Most states issued school closure rules and restaurant restrictions by April 7, 2020.

in the logarithmic scale. Notably, the propagation of Covid-19 is in exponential growth. Therefore, using the logarithmic scale curve makes it more accessible to detect the phase change of increase trend and determine the outbreak date accordingly [14]. In terms of Fig. 2, we can observe an apparent step-up around March 1, 2020, as denoted by the red point. Hence, we regard March 1, 2020, as the outbreak date of Covid-19 in North America.

We then begin the analysis by comparing the distributions of smartphone usage variables before and after the outbreak of Covid-19. In Fig. 3, we use box-plots to depict the distributions of the percentages of CPU usage, memory usage, WiFi usage, and network switches, respectively. Specifically, the 'Before' set contains the samples from November 1, 2019, to February 29, 2020, while the 'After' set contains the samples from March 1, 2020, to April 30, 2020. The box-plots describe data distribution through quartiles. The candlesticks represent the minimum and the maximum values, while the boxed area contains the values between 25% and 75% quartiles. The horizontal line denotes the median, while the green upper triangle indicates the mean.

There is an apparent difference in smartphone usage across the outbreak in terms of all hardware variables. The mean values of CPU and memory usage drop from 7.36% and 3.93% to 6.87% and 3.47%, respectively. Their differences across the outbreak are significant under a two-sided t -test [15] with p values of $5.239 \cdot 10^{-5} \ll 0.001$ and $7.383 \cdot 10^{-18} \ll 0.001$. The decreases imply that users' smartphone engagement becomes less active after the outbreak, i.e., March 1, 2020. Meanwhile, the WiFi usage percentage grows dramatically, where the mean value rises from 56.95% to 64.06%. The distribution difference is also significant under a two-sided t -test with a p value of $2.585 \cdot 10^{-19} \ll 0.001$. Since WiFi access points are usually deployed indoors, we can conclude that people have more time to stay indoors instead of going outside after the outbreak of Covid-19. Moreover, we also notice that the percentage of network switches drops remarkably. The mean value declines from 3.98% to 2.85%, and the distribution difference is significant, with p value $1.526 \cdot 10^{-23} \ll 0.001$. Similar to

³The code is available at <https://github.com/carat-project/carat/>.

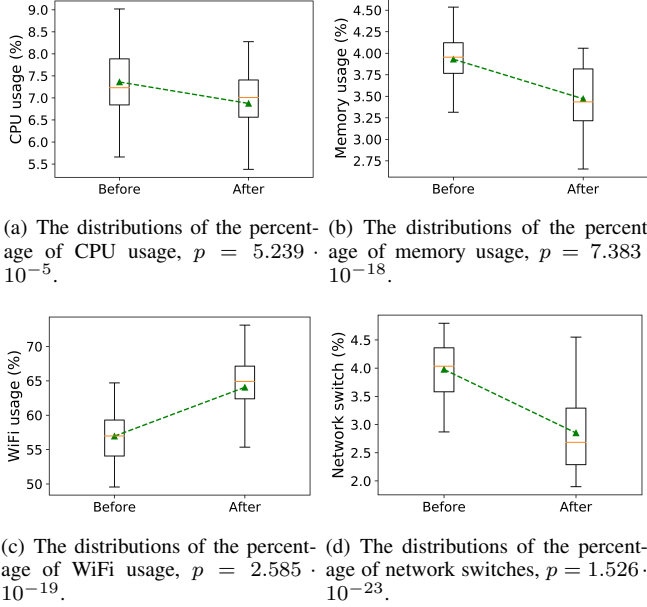


Fig. 3. The differences in smartphone usage before and after the outbreak of Covid-19.

TABLE III
CORRELATIONS BETWEEN MEMORY USAGE AND WiFi USAGE ACROSS DIFFERENT TIME PERIODS.

Over complete time window	Before outbreak	After outbreak
-0.0689	0.3205	-0.3240

WiFi usage, network switches also reflect the movement of mobile users. Since the WiFi network is commonly deployed indoors and limited by its coverage, network switches usually occur when mobile users go from indoors to outside and from outside to indoors. Consequently, the percentage of network switches can reveal the mobility intensity of smartphone users. In this way, the decreasing trend of network switches suggests users have less mobility after the outbreak.

As a result, based on the differences in number and distributions, we can conclude that the outbreak of Covid-19 causes a decrease in smartphone engagement in terms of both CPU and memory usage. Meanwhile, the outbreak causes an increase in users' intensity staying indoors in terms of WiFi usage. Further, we depict the correlation between WiFi usage and memory usage to investigate smartphone usage intensity when people stay indoors. Table III shows the correlations across different time periods, i.e., over the complete time windows, before the outbreak, and after the outbreak. As depicted in Table III, WiFi usage and memory usage have a weak positive correlation before the outbreak, which follows the commonly-held intuition. However, after the outbreak, the correlation becomes weak negative. We infer that the longer time to stay at home after the outbreak may cause such differences. When people have more time at home, they will prefer to use their computers and laptops for entertainment instead of smartphones.

B. Differences in Diurnal Patterns

In terms of the above statistical analysis, we can conclude that the outbreak of Covid-19 has affected users' smartphone usage behavior. Next, we delve into the dynamic analysis, i.e., revealing the differences in diurnal patterns. The diurnal pattern depicts how users' smartphone usage behavior unfolds over the time of the day, which is an essential temporal pattern studied by many previous studies [16], [17].

We define each day's diurnal pattern by averaging the usage data over the day's active users. In our case, we evenly divide one day into 48 time-slots, where each time-slot represents half an hour. Therefore, each diurnal sequence is of 48 dimensions. Next, we compute smartphone usage data for each time-slot. In practice, as for CPU usage and memory usage behavior, we take the averages in that time slot. For WiFi usage, we calculate the proportion of WiFi connection records in that time slot. Besides, for network switches, we calculate the proportion of network type changes in the time slot. By doing so, given one day, each type of smartphone usage behavior will have a diurnal sequence with 48 dimensions. In total, we have 728 diurnal sequences, i.e., $182 (\# \text{ of days}) \times 4 (\# \text{ of usage types})$.

After obtaining the diurnal sequences, we use the t-SNE transformation [18] to visualize them, as shown in Fig. 4. t-SNE is a commonly used data transformation method that projects high-dimensional data to a low-dimensional space while keeping the similarity across objects. In Fig. 4, blue points represent the dates before the outbreak, while orange points represent the dates after the outbreak. We can observe that excluding CPU usage, the other types of smartphone usage behavior appear to be nicely separated by the outbreak. This shows the existence of differences in diurnal patterns of smartphone usage before and after the outbreak of Covid-19.

Based on the t-SNE visualization results, we propose a hypothesis that the outbreak of Covid-19 will lead to a new diurnal pattern for smartphone usage. In our case, the new pattern means that it does not or rarely appears before the outbreak but is popular on the dates after the outbreak. To test the hypothesis, we apply K-means to cluster diurnal sequences of the entire 182 days for each type of smartphone usage behavior and examine whether the cluster results can be distinguished by the outbreak date of Covid-19. Since there are only two situations for any date, i.e., before or after the outbreak, we set the number of clusters to two. The clustering results are presented in Fig. 5~8, where the cluster A and B refer to the two-cluster output of K-means. Also, we regard the centroid as the typical diurnal pattern of the cluster.

Diurnal patterns of CPU usage. As shown in Fig. 5(a), the obtained two typical diurnal patterns of CPU usage have the same trend but different values. Both of them decrease during the night and increase during the day, while cluster B's centroid is of lower numerical values. Fig. 5(b) shows that, compared to cluster A, cluster B accounts for a higher proportion of the dates after the outbreak, consistent with the dropping trend observed in Fig. 3(a). We also observe that Covid-19 only affects the proportion of two cluster labels, and both typical patterns frequently appear on the dates before the

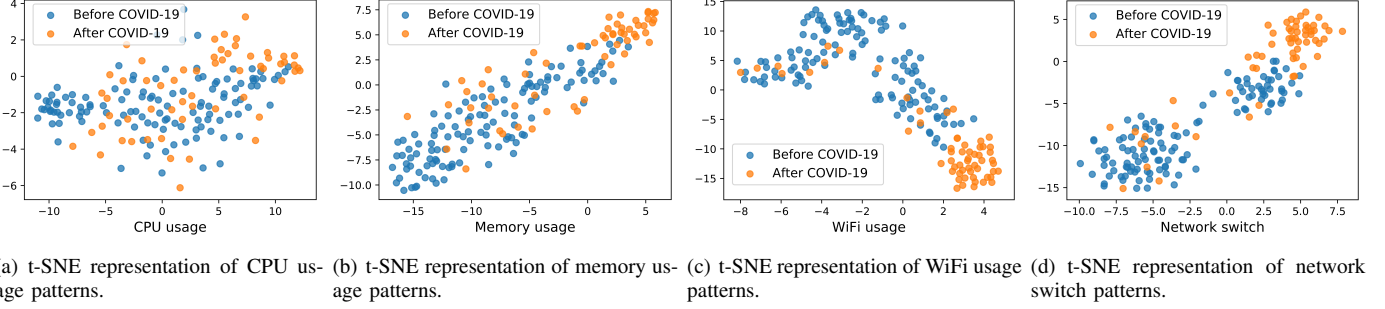


Fig. 4. t-SNE representation of diurnal sequences of smartphone usage, projecting high-dimensional data to a 2-dimensional space while keeping the similarity across objects.

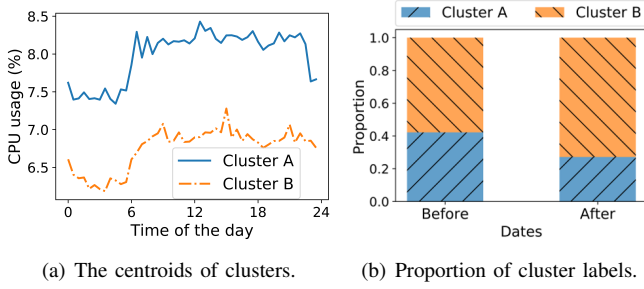


Fig. 5. Cluster results of CPU usage diurnal patterns.

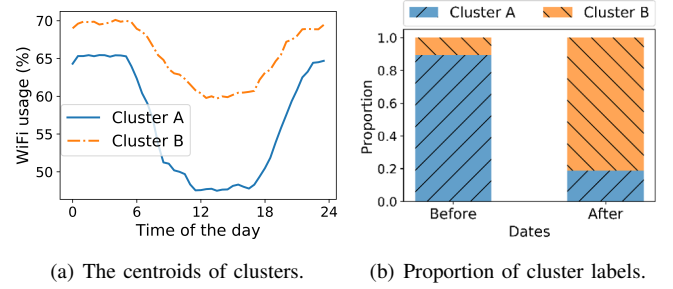


Fig. 7. Cluster results of WiFi usage diurnal patterns.

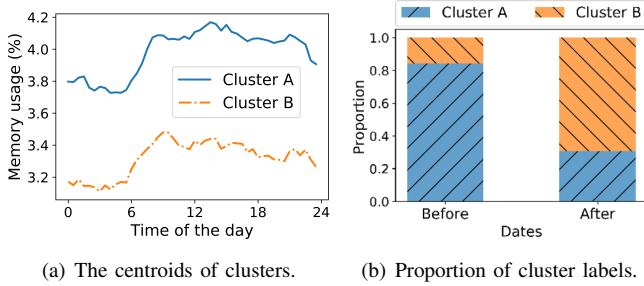


Fig. 6. Cluster results of memory usage diurnal patterns.

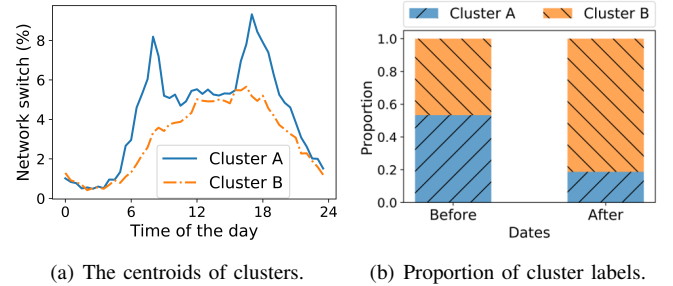


Fig. 8. Cluster results of network switch diurnal patterns.

outbreak. In other words, the outbreak did not create a new typical diurnal pattern of CPU usage. The t-SNE visualization in Fig. 4(a) also verifies this.

Diurnal patterns of memory usage. As depicted in Fig. 6(a), similar to CPU usage, two typical diurnal patterns obtained are also with the same trend but different numerical values. In terms of Fig. 6(b), over 80% of the dates before the outbreak belong to cluster A. Meanwhile, more than 65% of the dates after the outbreak belong to cluster B. Therefore, we can conclude that the cluster results can be distinguished by the outbreak date. Also, cluster B's centroid can be regarded as a new typical diurnal pattern because it rarely appears before the outbreak and becomes common after the outbreak. In summary, Covid-19 leads to the appearance of a new typical diurnal pattern of memory usage, corresponding to the t-SNE visualization in Fig. 4(b).

Diurnal patterns of WiFi usage. Fig. 7 displays the cluster results of WiFi usage. Unlike CPU and memory usage, apart from numerical differences, the centroids of WiFi usage

clusters also have different changing trends. As depicted in Fig. 7(a), the centroid of cluster B has a higher percentage of WiFi usage throughout the day. Instead of a cliff-like drop shown in cluster A, cluster B has a slow-down after 6 am. This indicates that users need less mobile network support on the dates in cluster B. Moreover, similar to memory usage, the dates after the outbreak have a dominating cluster, i.e., cluster B. Therefore, Covid-19 also brings a new diurnal pattern of WiFi usage, leading users to use more WiFi connections.

Diurnal patterns of network switches. We exhibit the clustering results of network switch patterns in Fig. 8. As discussed in Section III-A, network switches can reflect the mobility intensity of smartphone users. In Fig. 8(a), the centroid of cluster A presents two peaks in the morning and evening rush hours, which verifies the above discussion. We notice that less than 18% of the dates after the outbreak belong to cluster A, indicating that users' mobility intensity drops significantly. Alternatively, cluster B has fewer network switches throughout the day and without bimodal patterns, indicating that users

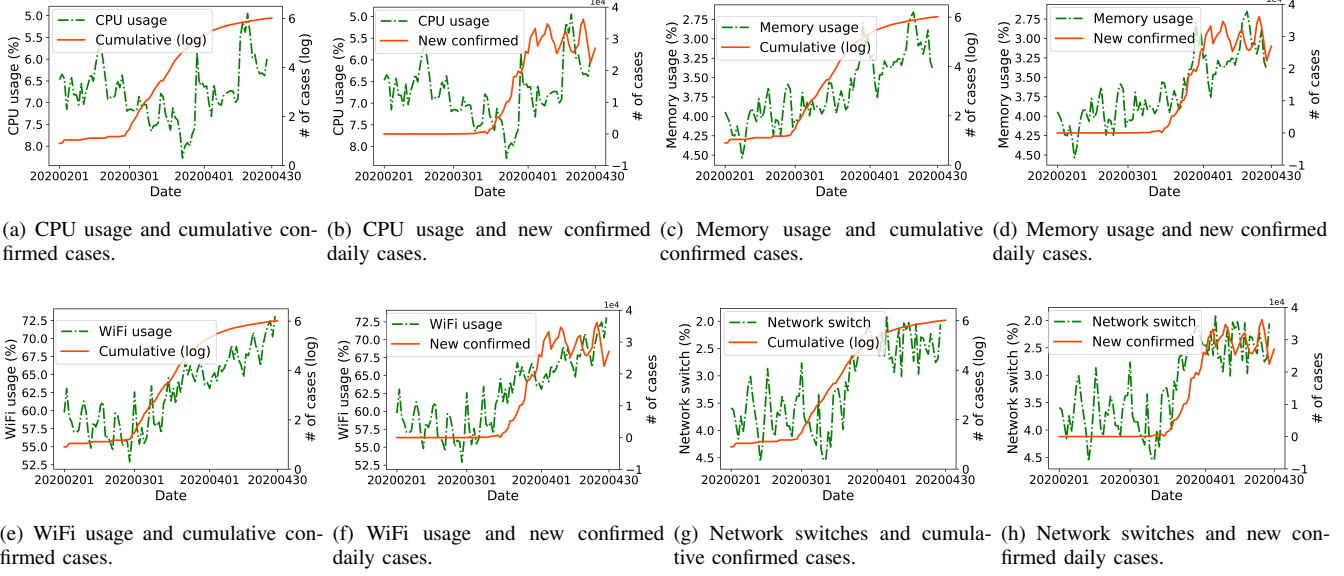


Fig. 9. The daily patterns of smartphone usage and the number of daily confirmed cases of Covid-19.

TABLE IV
PEARSON CORRELATIONS BETWEEN SMARTPHONE USAGE AND COVID-19 CASES.

Delay (day)	CPU usage		Memory usage		WiFi usage		Network switch	
	Cumulative (log)	New confirmed	Cumulative (log)	New confirmed	Cumulative (log)	New confirmed	Cumulative (log)	New confirmed
0	-0.0632	-0.2670	-0.8163	-0.8119	0.8364	0.8053	-0.7649	-0.7809
1	-0.0383	-0.2339	-0.8119	-0.8091	0.8754	0.8253	-0.7838	-0.8234
2	-0.0541	-0.2444	-0.8315	-0.8369	0.8667	0.8263	-0.7830	-0.8369
3	-0.0240	-0.2081	-0.8277	-0.8399	0.8523	0.8312	-0.7772	-0.8430

have less mobility on the dates in that cluster. Although cluster B dominates the dates after the outbreak, it also frequently appears before the outbreak. As a result, similar to CPU usage, Covid-19 only changes the proportion of different network switch patterns but does not trigger the appearance of new patterns.

Consequently, the outbreak of Covid-19 also profoundly affects diurnal patterns of smartphone usage behavior, implying that the diurnal sequences of smartphone usage can be used to reflect the outbreak status.

C. Correlations Between Smartphone Usage and Covid-19 Daily Cases

We then analyze the correlations between smartphone usage and Covid-19 daily cases. Specifically, we take the average over the active users of each day and plot both the daily sequences of smartphone usage and the number of daily confirmed cases of Covid-19 in Fig. 9, from February 1, 2020 to March 30, 2020. For the figures of CPU usage, memory usage, and network switches, we inverse the y-axis for better visualization. From the results, we can observe that memory usage, WiFi usage, and network switches have strong correlations with both cumulative and new confirmed daily cases. That is because smartphone usage behavior reflects users' physical activities, e.g., staying at home and mobility

intensity. Meanwhile, users' physical activities will influence and be affected by Covid-19. Therefore, smartphone usage behavior can indirectly reveal Covid-19 trends. Moreover, in Fig. 9(f) and Fig. 9(h), we discover a delay in the changing trends between smartphone usage behavior and new confirmed cases. In other words, smartphone usage changes earlier than the number of cases.

Further, to better explore the delay phenomenon, we put a set of delays on the daily sequences of smartphone usage behavior from 0 to 3 days. Then, we compute the Pearson correlation between shifted smartphone usage and Covid-19 sequences. The results are illustrated in Table IV. From the results, we discover that different smartphone usage features have various correlations with daily confirmed cases. Generally, memory usage, WiFi usage, and network switches have significant linear correlations with Covid-19 daily confirmed cases. The absolute values of their Pearson coefficients are greater than 0.8. However, CPU usage has a weak Pearson correlation, only around -0.26, with new confirmed cases of Covid-19. These observations are consistent with the findings in Section III-B. Moreover, when we delay usage behavior, it will have a higher correlation with Covid-19 cases, which corresponds to the observation that smartphone usage changes earlier than Covid-19 cases. Also, different smartphone usage variables show different typical time delays. In summary, the

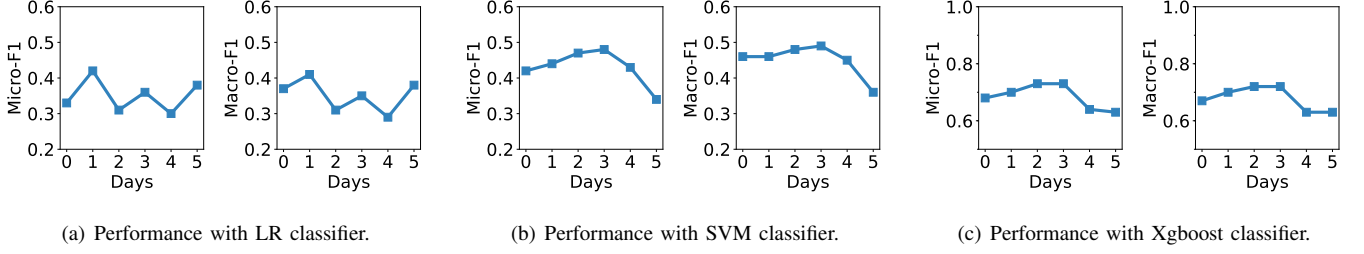


Fig. 10. Covid-19 outbreak stage inferences with different time delays.

correlations between smartphone usage behavior and daily confirmed cases present a high potential of using smartphone usage for daily outbreak stage inference of Covid-19.

IV. INFERENCE OF OUTBREAK STAGES

In this section, we study the second research problem, i.e., whether we can use smartphone usage data, e.g., CPU usage, memory usage, and network connections, to infer the outbreak stages of Covid-19. The outbreak stages reflect different severities of the pandemic. Specifically, we try to determine two points, i.e., the typical time delay of stage inference using smartphone usage data and the performance of different smartphone usage features in Covid-19 stage inference. Also, to further improve inference performance, we propose an embedding mechanism to fuse different smartphone usage behavior features.

A. Inference Settings

Recalling Fig. 2, we can witness that the outbreak of Covid-19 has shown three stages from March 1, 2020, to April 30, 2020. First, the dates from February 1, 2020, to March 1, 2020, are the early stage of Covid-19, with only a few cases appearing. Second, during the dates from March 1, 2020, to April 1, 2020, the daily confirmed cases increased dramatically. Third, on the dates after April 1, 2020, the increasing trend of Covid-19 cases is stable. Therefore, we label Covid-19 outbreak stages with three classes, i.e., early, dramatic, and stable. By doing so, the inference problem is converted into a 3-class classification problem. Specifically, we infer the outbreak stages of one day by using its diurnal sequences of different smartphone usage behavior, including CPU usage, memory usage, WiFi usage, and network switches. Also, to evaluate the performance, we use Macro-F1 and Micro-F1 as metrics. Macro-F1 treats all classes equally, computing the F1-score independently for each class and then taking the average. Alternatively, Micro-F1 aggregates the contributions of all classes to compute the average F1-score. The higher the value of Macro-F1 and Micro-F1, the better the performance. For all experiments, we obtain the results by employing a five-fold cross-validation policy on our dataset.

B. Delay Analysis of Stage Inference

As we have discussed in Section III, users' smartphone usage behavior can reflect their physical activities and the outbreak stages of Covid-19. However, the reflection may

not be immediately expressed by the daily cases of Covid-19 due to the incubation period and diagnosis delay. Hence, we explore the typical time delay of stage inference. Specifically, we infer the outbreak stage of one day by utilizing the smartphone usage features of the days before it. We use inference performance to evaluate the correlations between smartphone usage and Covid-19 trends. In other words, better performance indicates a higher correlation. Notably, different from the Pearson correlation, the task of inference can also reveal nonlinear correlations. In practice, we conduct the inference with the three most commonly used classification algorithms, logistic regression (LR) [19], support vector machine (SVM) [20] and Xgboost [21]. We infer the outbreak stages of one day by concatenating all behavior types' diurnal sequences, including CPU usage, memory usage, WiFi usage, and network switches.

We show the results in Fig. 10. The LR classifier has poor performance, and F1 scores fluctuate on different delays. That is because the LR classifier only uses a logistic function to model the correlation, which is more susceptible to outliers tampering with the performance. Therefore, it is hard to capture the relations between smartphone usage features and Covid-19 outbreak stages with the LR classifier using the real-world dataset that might have noisy data points. Alternatively, as shown in Fig. 10(b) and Fig. 10(c), SVM and Xgboost classifiers have better performance. Also, we can observe that F1 scores achieve the highest value under a delay of 2 or 3 days. This observation confirms that the reflection of users' smartphone usage behavior will emerge in Covid-19 trends with a time delay of a few days, further validating our analysis in Section III-C.

C. Performance of Different Usage Features

Next, we evaluate the performance of different smartphone usage features and their combinations for the Covid-19 outbreak stage inference. Specifically, we explore four types of smartphone usage features, i.e., CPU usage (CPU), memory usage (Mem), WiFi usage (WiFi), and network switches (Net). We combine a set of features by concatenating them together. We perform the inference with the Xgboost classifier. The performance of different combinations of features is shown in Table V.

For the inference with a single feature, the performance of using network switches is the best, indicating that users' mobility intensity is most relevant to the Covid-19 status. Meanwhile, WiFi and memory usage achieve relatively good performance, implying that WiFi and memory usage also

TABLE V
INFERENCE PERFORMANCE WITH DIFFERENT FEATURES.

Features	Macro-F1	Micro-F1
CPU	0.590	0.584
Mem	0.702	0.697
WiFi	0.713	0.708
Net	0.757	0.753
CPU+Mem	0.666	0.663
CPU+WiFi	0.722	0.719
CPU+Net	0.609	0.685
Mem+WiFi	0.683	0.674
Mem+Net	0.735	0.730
WiFi+Net	0.707	0.696
CPU+Mem+WiFi	0.716	0.708
CPU+WiFi+Net	0.715	0.707
Net+Mem+WiFi	0.739	0.730
CPU+Mem+Net	0.766	0.764
CPU+Mem+Net+WiFi	0.733	0.721

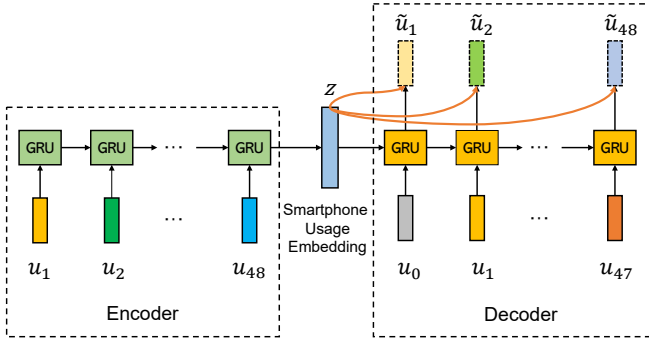


Fig. 11. Seq2Seq model for smartphone usage embedding.

reflect crucial human behavior related to Covid-19. In contrast, the CPU usage feature is less related and with the lowest inference performance. These inference results are consistent with our findings in Section III. As for the inference with multiple features, the performance is not simply a superposition of single features' performance. In terms of Table V, the best performance is achieved by using CPU, memory, and WiFi usage. However, it only achieves the F1 score of around 0.76, slightly higher than when merely using network switches. Also, most cases of using multiple features have lower performance than simply using network switches. These results reveal that simple concatenation is insufficient to fuse different behavior data, motivating us to develop a better fusion mechanism to explore different features effectively.

D. Smartphone Usage Behavior Embedding

In this section, we propose an embedding model to fuse different smartphone usage behavior effectively. Given a day, we first construct a diurnal smartphone usage feature sequence $\{u_i\}_{i=1}^{48}$, where u_i is a vector containing all four usage features in the i -th timeslot of the day. We then utilize a Seq2Seq [22] model to learn an embedding from the diurnal sequence. As shown in Fig. 11, the model consists of an encoder and a decoder, which are implemented with a GRU network [22]. The sequence $\{u_i\}_{i=1}^{48}$ is fed into the encoder to obtain an

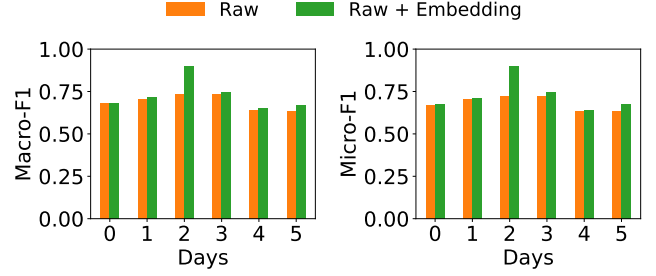


Fig. 12. Outbreak stage inferences with embeddings.

encoding vector of z . Then, z and a shifted usage sequence $\{u_i\}_{i=0}^{47}$ are fed into the decoder to reconstruct the original sequence, where u_0 is a vector that contains all 1. Moreover, to encode comprehensive information in vector z , we engage z in the reconstruction. Formally, the i -th unit of the decoder takes u_{i-1} as input and outputs hidden state \hat{h}_i , we infer \hat{u}_i as,

$$\hat{u}_i = \sigma(W[\hat{h}_i, z] + b), \quad (1)$$

where $[\cdot]$ is the concatenating operation, σ is the sigmoid activating function, W and b are trainable parameters. Finally, we train the model by minimizing the reconstruction loss,

$$\mathcal{L} = \sum_{i=1}^{48} |\hat{u}_i - u_i|^2. \quad (2)$$

In our experiment, we train the model with the Adam optimizer with a learning rate of 0.0001. The batch size is set as the number of sequences, and we train the model for 200 epochs. By doing so, we obtain a usage embedding vector for each day. To evaluate whether the embedding fuses different usage features better, we conduct the inference on the original features (Raw) and the original features concatenated with the learned embeddings (Raw + Embedding). We again use the Xgboost classifier as the inference model.

We compare the performance with embeddings, as shown in Fig. 12. We can observe that, by combining with embeddings, we improve the entire performance under different delay settings. Especially when the delay is set as two days, the performance of raw features combined with embeddings reaches around 0.87 for both Macro-F1 and Micro-F1, which has an over 20% improvement compared with the best performance of only using raw features. These results demonstrate that the learned embeddings fuse multiple features more effectively indeed.

V. DISCUSSION AND LIMITATION

In this paper, we have investigated the impact of Covid-19 on smartphone usage based on a real-world dataset. However, the number of users involved in our dataset is not very large, i.e., covering 425 users, which is a limitation of our work. The limited number of users involved may threaten the representativeness of our conclusion. To alleviate the influence caused by a limited number of users, we have taken several adequate measures in our work. For example, we compared the distribution of variables instead of the average and median. We also used the p-value to verify statistical significance.

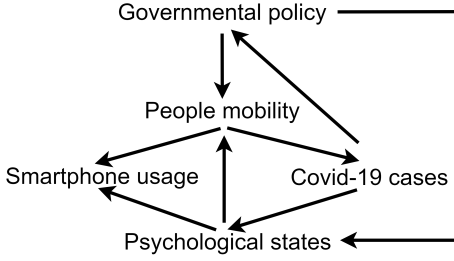


Fig. 13. A potential causality diagram of smartphone usage and Covid-19 cases.

Although we have examined the correlation between smartphone usage behavior and Covid-19 cases, their causality relationship still needs further exploration. In Fig. 13, we depict a potential causality diagram of smartphone usage and Covid-19 cases. People mobility and psychological state serve as a confounder and mediator connecting smartphone usage and Covid-19 cases, respectively. Smartphone usage is directly affected by mobility and can act as a mobility indicator. Also, smartphone usage is still affected by the psychological states of users [13]. Meanwhile, the causation between people mobility and Covid-19 cases is bidirectional. On the one hand, frequent people mobility will trigger new Covid-19 cases. On the other hand, Covid-19 will affect people's mobility through governmental policies and their psychological states. Therefore, the causation between smartphone usage and Covid-19 cases might be complex. As for checking the potential causality diagram we proposed, we leave it to future work.

VI. RELATED WORK

Many previous studies have focused on characterizing smartphone usage behavior. Shafiq *et al.* [23] presented the diurnal pattern of smartphone network usage from various granularities, i.e., bytes, packets, flows, and users. Peltonen *et al.* [24] collected a one-year smartphone usage dataset from 25,323 users distributed in 44 countries. They then studied how cultural features affect users' smartphone usage behavior. Srinivasan *et al.* [25] indicated that smartphone usage behavior profoundly depends on contextual information. For example, users use more WiFi connections at home. Moreover, Van Canneyt *et al.* [26] exhibited that the occurrence of special events, e.g., New year's day, UEFA European Championship, will disrupt users' normal smartphone usage patterns. These existing studies demonstrate that users' smartphone usage behavior will be sensitively impacted by diverse contextual factors, including time, locations, and big events, which inspired us to investigate how the outbreak of Covid-19 affects the smartphone usage behavior.

Also, some studies pointed out the strong link between smartphone usage behavior and users' physical attributes and activities. Zhao *et al.* [27] analyzed one month of smartphone usage data collected from 106,762 users. They then discovered 382 distinct types of users based on their usage behavior. Also, they gave each cluster a meaningful label, such as night communicators, evening learners, and financial users. Do *et al.* [28] represented users' smartphone usage traces in one day as a bag-of-words, where one word refers to a smartphone

usage record with time features. They then applied an author-topic model to infer the underlying structure of users' physical activities. Similarly, Li *et al.* [29] leveraged smartphone app usage data to identify users' daily activities. These studies demonstrated that users' physical activities profoundly shape smartphone usage behavior, which shed light on using smartphone usage data to reflect human activities and further infer Covid-19 outbreak stages.

Smartphone usage behavior is still affected by users' psychological states. For example, Saeb *et al.* [30] explored smartphone sensors' data, like accelerometer, screen, GPS, and WiFi, which help estimate the depression and anxiety of users. Their methods can also be applied to our dataset, allowing us to detect the depression of users. During the Covid-19 crisis, we need to pay more attention to mental health in the population. Covid-19 may trigger psychiatric disorders of people [31]. Elhai *et al.* analyzed gaming disorder severity [32] and anxiety symptoms [33] during Covid-19. Moreover, Montag *et al.* [13] pointed out that we can leverage smartphone data to detect population mental states in real-time to help fight the Covid-19 pandemic. They also developed an app [34] for social scientists, which tracks smartphone usage data by combining self-report data with objectively recorded data. In practice, conducting population-scale digital phenotyping might be challenging due to the lack of sufficient labeled data. In that case, label-less learning should be a helpful technology. For example, Chen *et al.* [35] proposed a label-less learning for emotion cognition on a large-scale.

Some studies also analyzed physical activities during Covid-19 by using smartphone app usage and sensory data. Norbury *et al.* [36] discovered a positive relation between social app usage and total footsteps (obtained from sensory data) during the lockdown due to Covid-19. Couture *et al.* [37] investigated county-to-county movements based on the GPS data collected from smartphones. Unlike the above studies, our work directly investigates the relation between smartphone usage and the Covid-19 outbreak.

VII. CONCLUSION

We conduct the first comprehensive study of the impact of Covid-19 on smartphone usage. Specifically, our analysis covers the mobile users in North America with six-month smartphone usage records from November 2019 to April 2020. Overall, our findings indicate that users' smartphone usage indeed changes across the outbreak of Covid-19. However, the outbreak has different effects on different usage behavior in terms of changing trends, diurnal patterns, and correlations. Also, we demonstrate the potential of using smartphone usage data to infer the outbreak stages, achieving over 0.8 for both Macro-F1 and Micro-F1. Our findings provide a novel application of smartphone usage data and explore their values for fighting against the epidemic.

ACKNOWLEDGMENTS

This research has been supported in part by project 16214817 from the Research Grants Council of Hong Kong, project FP805 from HKUST, the 5GEAR project, the FIT

project and the CBAI (Crowdsourced Battery Optimization AI for a Connected World, grant No. 1319017) project from the Academy of Finland, the National Natural Science Foundation of China under U1936217, 61971267, 61972223, 61941117, 61861136003, Beijing Natural Science Foundation under L182038, Beijing National Research Center for Information Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology.

REFERENCES

- [1] T. P. Velavan and C. G. Meyer, "The covid-19 epidemic," *Tropical medicine & international health*, vol. 25, no. 3, p. 278, 2020.
- [2] W. McKibbin and R. Fernando, "The economic impact of covid-19," *Economics in the Time of COVID-19*, vol. 45, 2020.
- [3] S. Li, Y. Wang, J. Xue, N. Zhao, and T. Zhu, "The impact of covid-19 epidemic declaration on psychological consequences: a study on active weibo users," *International journal of environmental research and public health*, vol. 17, no. 6, p. 2032, 2020.
- [4] M. Chen, M. Li, Y. Hao, Z. Liu, L. Hu, and L. Wang, "The introduction of population migration to sear for covid-19 epidemic modeling with an efficient intervention strategy," *Information Fusion*, vol. 64, pp. 252–258, 2020.
- [5] T. Li, M. Zhang, H. Cao, Y. Li, S. Tarkoma, and P. Hui, "“” what apps did you use?”: Understanding the long-term evolution of mobile app usage," in *Proceedings of The Web Conference 2020*, 2020, pp. 66–76.
- [6] Statista, "https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/," 2020.
- [7] M. Chen, Y. Jiang, N. Guizani, J. Zhou, G. Tao, J. Yin, and K. Hwang, "Living with i-fabric: smart living powered by intelligent fabric and deep analytics," *IEEE Network*, vol. 34, no. 5, pp. 156–163, 2020.
- [8] T. Yarkoni, "Psychoinformatics: New horizons at the interface of the psychological and computing sciences," *Current Directions in Psychological Science*, vol. 21, no. 6, pp. 391–397, 2012.
- [9] T. R. Insel, "Digital phenotyping: technology for a new science of behavior," *Jama*, vol. 318, no. 13, pp. 1215–1216, 2017.
- [10] H. Baumeister and C. Montag, *Digital Phenotyping and Mobile Sensing*. Springer, 2019.
- [11] A. Markowetz, K. Błaskiewicz, C. Montag, C. Switala, and T. E. Schlaepfer, "Psycho-informatics: big data shaping modern psychometrics," *Medical hypotheses*, vol. 82, no. 4, pp. 405–411, 2014.
- [12] J. J. Van Bavel, K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M. J. Crockett, A. J. Crum, K. M. Douglas, J. N. Druckman *et al.*, "Using social and behavioural science to support covid-19 pandemic response," *Nature human behaviour*, vol. 4, no. 5, pp. 460–471, 2020.
- [13] C. Montag, P. Dagum, J. D. Elhai *et al.*, "On the need for digital phenotyping to obtain insights into mental states in the covid-19 pandemic," *Digital Psychology*, vol. 1, no. 2, pp. 40–42, 2020.
- [14] B. F. Maier and D. Brockmann, "Effective containment explains subexponential growth in recent confirmed covid-19 cases in china," *Science*, vol. 368, no. 6492, pp. 742–746, 2020.
- [15] J. Neyman and E. S. Pearson, "The testing of statistical hypotheses in relation to probabilities a priori," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 29, no. 4. Cambridge University Press, 1933, pp. 492–510.
- [16] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Characterizing geospatial dynamics of application usage in a 3g cellular data network," in *2012 Proceedings IEEE INFOCOM*. IEEE, 2012, pp. 1341–1349.
- [17] N. Ghahramani and C. Brakewood, "Trends in mobile transit information utilization: An exploratory analysis of transit app in new york city," *Journal of Public Transportation*, vol. 19, no. 3, p. 9, 2016.
- [18] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
- [19] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.
- [20] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [21] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [22] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
- [23] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Geospatial and temporal dynamics of application usage in cellular data networks," *IEEE Transactions on mobile computing*, vol. 14, no. 7, pp. 1369–1381, 2014.
- [24] E. Peltonen, E. Lagerspetz, J. Hamberg, A. Mehrotra, M. Musolesi, P. Nurmi, and S. Tarkoma, "The hidden image of mobile apps: Geographic, demographic, and cultural factors in mobile usage," in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2018, pp. 1–12.
- [25] V. Srinivasan, S. Moghaddam, A. Mukherji, K. K. Rachuri, C. Xu, and E. M. Tapia, "Mobileminer: Mining your frequent patterns on your phone," in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, 2014, pp. 389–400.
- [26] S. Van Canneyt, M. Bron, A. Haines, and M. Lalmas, "Describing patterns and disruptions in large scale mobile app usage data," in *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017, pp. 1579–1584.
- [27] S. Zhao, J. Ramos, J. Tao, Z. Jiang, S. Li, Z. Wu, G. Pan, and A. K. Dey, "Discovering different kinds of smartphone users through their application usage behaviors," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 498–509.
- [28] T.-M.-T. Do and D. Gatica-Perez, "By their apps you shall understand them: mining large-scale patterns of mobile phone usage," in *Proceedings of the 9th international conference on mobile and ubiquitous multimedia*, 2010, pp. 1–10.
- [29] T. Li, Y. Li, M. A. Hoque, T. Xia, S. Tarkoma, and P. Hui, "To what extent we repeat ourselves? discovering daily activity patterns across mobile app usage," *IEEE Transactions on Mobile Computing*, 2020.
- [30] S. Saeb, E. G. Lattie, K. P. Kording, and D. C. Mohr, "Mobile phone detection of semantic location and its relationship to depression and anxiety," *JMIR mHealth and uHealth*, vol. 5, no. 8, p. e112, 2017.
- [31] A. Schimmenti, J. Billieux, and V. Starcevic, "The four horsemen of fear: An integrated model of understanding fear experiences during the covid-19 pandemic," *Clinical Neuropsychiatry*, vol. 17, no. 2, pp. 41–45, 2020.
- [32] J. D. Elhai, D. McKay, H. Yang, C. Minaya, C. Montag, and G. J. Asmundson, "Health anxiety related to problematic smartphone use and gaming disorder severity during covid-19: Fear of missing out as a mediator," *Human Behavior and Emerging Technologies*, vol. 3, no. 1, pp. 137–146, 2021.
- [33] J. D. Elhai, H. Yang, D. McKay, and G. J. Asmundson, "Covid-19 anxiety symptoms associated with problematic smartphone use severity in chinese adults," *Journal of Affective Disorders*, vol. 274, pp. 576–582, 2020.
- [34] C. Montag, H. Baumeister, C. Kannen, R. Sariyska, E.-M. Meßner, and M. Brand, "Concept, possibilities and pilot-testing of a new smartphone application for the social and life sciences to study human behavior including validation data from personality psychology," *J—Multidisciplinary Scientific Journal*, vol. 2, no. 2, pp. 102–115, 2019.
- [35] M. Chen and Y. Hao, "Label-less learning for emotion cognition," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 7, pp. 2430–2440, 2019.
- [36] A. Norbury, S. H. Liu, J. J. Campaña-Montes, L. Romero-Medrano, M. L. Barrigón, E. Smith, A. Artés-Rodríguez, E. Baca-García, and M. M. Perez-Rodríguez, "Social media and smartphone app use predicts maintenance of physical activity during covid-19 enforced isolation in psychiatric outpatients," *Molecular psychiatry*, pp. 1–11, 2020.
- [37] V. Couture, J. I. Dingel, A. E. Green, J. Handbury, and K. R. Williams, "Measuring movement and social contact with smartphone data: a real-time application to covid-19," National Bureau of Economic Research, Tech. Rep., 2020.



Tong Li received the B.S. degree and M.S. degree in communication engineering from Hunan University, China, in 2014 and 2017. At present, he is a dual Ph.D. student at the Hong Kong University of Science and Technology and the University of Helsinki. His research interests include data mining and machine learning, especially with applications to mobile big data and urban computing. He is an IEEE student member.



Mingyang Zhang is currently pursuing the Ph.D. degree with the department of Computer Science and Engineering, Hong Kong University of Science and Technology, within the System and Media Laboratory (SymLab). He received the B.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 2018. His research interests include spatiotemporal data mining, urban computing.



Yong Li (M'09-SM'16) is currently a Tenured Associate Professor of the Department of Electronic Engineering, Tsinghua University. He received the Ph.D. degree in electronic engineering from Tsinghua University in 2012. His research interests include machine learning and big data mining, particularly, automatic machine learning and spatial-temporal data mining for urban computing, recommender systems, and knowledge graphs. Dr. Li has served as General Chair, TPC Chair, SPC/TPC Member for several international workshops and

conferences, and he is on the editorial board of two IEEE journals. He has published over 100 papers on first-tier international conferences and journals, including KDD, WWW, UbiComp, SIGIR, AAAI, TKDE, TMC etc, and his papers have total citations more than 8300. Among them, ten are ESI Highly Cited Papers in Computer Science, and five receive conference Best Paper (run-up) Awards. He received IEEE 2016 ComSoc Asia-Pacific Outstanding Young Researchers, Young Talent Program of China Association for Science and Technology, and the National Youth Talent Support Program.



Emil Lagerspetz is a Postdoctoral Researcher at the Department of Computer Science, University of Helsinki. He completed his Ph.D. in Computer Science at the University of Helsinki in 2014. His research interests include large-scale data analysis (Big Data), edge computing, ubiquitous computing, and energy efficiency.



Sasu Tarkoma (SM'12) received the MSc and PhD degrees in computer science from the Department of Computer Science, University of Helsinki. He is a Professor of Computer Science at the University of Helsinki, and Head of the Department of Computer Science. He has authored 4 textbooks and has published over 200 scientific articles. His research interests are Internet technology, distributed systems, data analytics, and mobile and ubiquitous computing. He is Fellow of IET and EAI. He has nine granted US Patents. His research has received

several Best Paper awards and mentions, for example at IEEE PerCom, IEEE ICDCS, ACM CCR, and ACM OSR.



Pan Hui (SM'14-F'18) received his Ph.D. degree from the Computer Laboratory at University of Cambridge, and both his Bachelor and MPhil degrees from the University of Hong Kong.

He is the Nokia Chair Professor in Data Science and Professor of Computer Science at the University of Helsinki. He is also the director of the HKUST-DT Systems and Media Lab at the Hong Kong University of Science and Technology. He was a senior research scientist and then a Distinguished Scientist for Telekom Innovation Laboratories (T-labs) Germany and an adjunct Professor of social computing and networking at Aalto University. His industrial profile also includes his research at Intel Research Cambridge and Thomson Research Paris. He has published more than 300 research papers and with over 17,500 citations. He has 30 granted and filed European and US patents in the areas of augmented reality, data science, and mobile computing. He has been serving on the organising and technical program committee of numerous top international conferences including ACM SIGCOMM, MobiSys, IEEE Infocom, ICNP, SECON, IJCAI, AAAI, ICWSM and WWW. He is an associate editor for the leading journals IEEE Transactions on Mobile Computing and IEEE Transactions on Cloud Computing. He is an IEEE Fellow, an ACM Distinguished Scientist, and a member of the Academia Europaea.